

CLAIMS

1. A method of processing data in a computer system comprising at least one host and at least one content addressable storage system which stores units of data for the at least one host, wherein the at least one host accesses a unit of data using a content address based at least in part on the content of the unit of data, the method comprising an act of:
 - (a) creating the content addresses for units of data to comprise first information that provides an indication of which units of data are written to the storage system proximate in time.
2. The method of claim 1, further comprising acts of:
 - (b) accessing one of the units of data by providing the content address for the one of the units of data to the storage system.
 - (c) considering the first information the content address for the one of the units of data in determining where within the storage system to store the one of the units of data.
3. The method of claim 1, wherein the act (a) comprises, for one of the units of data, acts of:
 - creating a hash of at least a portion of the unit of data; and
 - inserting the first information in the hash to create the content address for the one of the units of data.
4. The method of claim 1, wherein the act of inserting the first information in the hash further comprises creating the first information at the at least one host.
5. The method of claim 1, wherein the act of inserting the first information in the hash further comprises creating the first information at the at least one storage system.
6. The method of claim 1, wherein the act of inserting the first information in the hash further comprises creating the first information at the at least one storage system and the at least one host.

7. The method of claim 3, wherein the first information includes at least a portion of a timestamp relating to when the one of the units of data is written to the at least one storage system.
8. The method of claim 2, wherein the act (b) comprises an act of using at least a portion of the first information to select a storage location within the at least one storage system for storing the one of the units of data, and wherein the method further comprises an act of storing the one of the units of data in the selected storage location.
9. The method of claim 2, wherein the first information is used by the at least one content addressable storage system to select a storage location within the at least one content addressable storage system for storing at least one of the units of data and wherein the storage location is a logical storage location within a file system on the at least one storage system.
10. The method of claim 2, wherein the method further comprises using the first information as a key to a database table.
11. The method of claim 2, wherein the first information is used by the at least one content addressable storage system to select a storage location within the at least one content addressable storage system for storing at least one of the units of data and wherein the storage location is a physical storage location within the at least one storage system.
12. At least one computer readable medium encoded with instructions that, when executed on a computer system perform a method of processing data, wherein the computer system comprises at least one host and at least one content addressable storage system which stores units of data for the at least one host, and wherein the at least one host accesses a unit of data using a content address based at least in part on the content of the unit of data, the method comprising an act of:

(a) creating the content addresses for units of data to comprise first information that provides an indication of which units of data are written to the storage system proximate in time.

13. The at least one computer readable medium of claim 12, wherein the method further comprises acts of:

(b) accessing one of the units of data by providing the content address for the one of the units of data to the storage system.

(c) considering the first information the content address for the one of the units of data in determining where within the storage system to store the one of the units of data.

14. The at least one computer readable medium of claim 13, wherein the act (a) comprises, for one of the units of data, acts of:

creating a hash of at least a portion of the unit of data; and

inserting the first information in the hash to create the content address for the one of the units of data.

15. The at least one computer readable medium of claim 14, wherein the act of inserting the first information in the hash further comprises creating the first information at the at least one host.

16. The at least one computer readable medium of claim 14, wherein the act of inserting the first information in the hash further comprises creating the first information at the at least one storage system.

17. The at least one computer readable medium of claim 14, wherein the act of inserting the first information in the hash further comprises creating the first information at the at least one storage system and the at least one host.

18. The at least one computer readable medium of claim 14, wherein the first information includes at least a portion of a timestamp relating to when the one of the units of data is written to the at least one storage system.

19. The at least one computer readable medium of claim 13, wherein the act (b) comprises an act of using at least a portion of the first information to select a storage location within the at least one storage system for storing the one of the units of data, and wherein the method further comprises an act of storing the one of the units of data in the selected storage location.

20. The at least one computer readable medium of claim 13, wherein the first information is used by the at least one content addressable storage system to select a storage location within the at least one content addressable storage system for storing at least one of the units of data and wherein the storage location is a logical storage location within a file system on the at least one storage system.

21. The at least one computer readable medium of claim 13, wherein the method further comprises using the first information as a key to a database table.

22. The at least one computer readable medium of claim 13, wherein the first information is used by the at least one content addressable storage system to select a storage location within the at least one content addressable storage system for storing at least one of the units of data and wherein the storage location is a physical storage location within the at least one storage system.

23. A host computer for use in a computer system comprising the host computer and a content addressable storage system, wherein the host computer accesses the data on the content addressable storage system using content addresses generated based on the content of the data units, the host computer comprising:

at least one processor that generates the units of data; and

at least one controller that creates the content addresses for units of data to comprise first information that provides an indication of which units of data are written to the content addressable storage system proximate in time.

24. The host computer of claim 23, wherein the controller accesses one of the units of data by providing the content address for the one of the units of data to the content addressable storage system.

25. The host computer of claim 24, wherein the controller, for one of the units of data:

- creates a hash of at least a portion of the unit of data; and
- inserts the first information in the hash to create the content address for the one of the units of data.

26. The host computer of claim 25, wherein the first information includes at least a portion of a timestamp relating to when the one of the units of data is written to the at least one content addressable storage system.

27. A method of processing data in a computer system comprising at least one host and at least one content addressable storage system which stores units of data for the at least one host, wherein the at least one host accesses a unit of data using a content address based at least in part on the content of the unit of data, wherein the at least one storage system comprises an index that maps content addresses for units of data to storage locations within the at least one storage system where the units of data are stored, the method comprising acts of:

- (a) receiving, from the at least one host, a request to store one of the units of data;

and

- (b) storing the one of the units of data in a storage location selected so that an entry in the index for the storage location is proximate to entries in the index for other units of data written to the at least one storage system proximate in time to the one of the units of data.

28. The method of claim 27, wherein the content address for the one of the units of data is created, at least in part, by hashing at least a portion of the unit of data.

29. The method of claim 28, wherein the content address includes first information that can be used to determine which units of data are written to the storage system proximate in time, and wherein the method further comprises receiving, from the at least one host, the first information.

30. The method of claim 28, further comprising an act of generating, at the at least one storage system, the first information that provides an indication of which units of data are written to the storage system proximate in time and adding the first information to the content address.

31. The method of claim 28, wherein the content address includes first information that can be used to determine which units of data are written to the storage system proximate and the method further comprises acts of:

generating a first portion of the first information at the at least one storage system;

receiving a second portion of the first information from the at least one host; and
adding the first portion and the second portion to the content address.

32. The method of claim 27, wherein the index is a file system having a plurality of directories, and wherein the act (b) comprises an act of storing the one of the units of data in a same directory as other units of data written to the at least one storage system proximate in time to the one of the units of data.

33. The method of claim 27, wherein the index is a database table, and wherein the act (b) comprises an act of storing the one of the units of data in a physical storage location proximate to the physical storage locations of other units of data written to the at least one storage system proximate in time to the one of the units of data.

34. At least one computer readable medium encoded with instructions that, when executed on a computer system, perform a method of processing data, wherein the computer system comprises at least one host and at least one content addressable storage system which stores units of data for the at least one host, wherein the at least one host

accesses a unit of data using a content address based at least in part on the content of the unit of data, and wherein the at least one storage system comprises an index that maps content addresses for units of data to storage locations within the at least one storage system where the units of data are stored, the method comprising acts of:

(a) receiving, from the at least one host, a request to store one of the units of data, the request identifying the one of the units of data with its associated content address; and

(b) storing the one of the units of data in a storage location selected so that an entry in the index for the storage location is proximate to entries in the index for other units of data written to the at least one storage system proximate in time to the one of the units of data.

35. The at least one computer readable medium of claim 34, wherein the content address for the one of the units of data is created, at least in part, by hashing at least a portion of the unit of data.

36. The at least one computer readable medium of claim 35, wherein the content address includes first information that can be used to determine which units of data are written to the storage system proximate in time, and wherein the method further comprises receiving, from the at least one host, the first information.

37. The at least one computer readable medium of claim 35, wherein the method further comprises an act of generating, at the at least one storage system, the first information that provides an indication of which units of data are written to the storage system proximate in time and adding the first information to the content address.

38. The at least one computer readable medium of claim 35, wherein the content address includes first information that can be used to determine which units of data are written to the storage system proximate and the method further comprises acts of:

generating a first portion of the first information at the at least one storage system;

receiving a second portion of the first information from the at least one host; and

adding the first portion and the second portion to the content address.

39. The at least one computer readable medium of claim 34, wherein the index is a file system having a plurality of directories, and wherein the act (b) comprises an act of storing the one of the units of data in a same directory as other units of data written to the at least one storage system proximate in time to the one of the units of data.

40. The computer readable medium of claim 34, wherein the index is a database table, and wherein the act (b) comprises an act of storing the one of the units of data in a physical storage location proximate to the physical storage locations of other units of data written to the at least one storage system proximate in time to the one of the units of data.

41. A content addressable storage system for use in a computer system that comprises the content addressable storage system and at least one host, wherein the at least one host accesses data units stored on the content addressable storage system using content addresses generated based on the content of the data units, the content addressable storage system comprising:

at least one storage device to store data received from the at least one host; and

at least one controller that maintains an index that maps content addresses for units of data to storage locations within the at least one storage system where the units of data are stored so that units of data written to the at least one storage system proximate in time have proximate entries in the index.

42. The content addressable storage system of claim 41, wherein the content addresses for the units of data are created, at least in part, by hashing at least a portion of the unit of data.

43. The content addressable storage system of claim 42, wherein the content address includes first information that can be used to determine which units of data are written to the storage system proximate in time, and wherein the at least one controller receives, from the at least one host, the first information.

44. The content addressable storage system of claim 42, wherein the controller generates the first information that provides an indication of which units of data are written to the storage system proximate in time and adds the first information to the content address.

45. The content addressable storage system of claim 42, wherein the content address includes first information that can be used to determine which units of data are written to the storage system proximate and the at least one controller:

generates a first portion of the first information at the at least one storage system;
receives a second portion of the first information from the at least one host; and
adds the first portion and the second portion to the content address.

46. The content addressable storage system of claim 41, wherein the index is a file system having a plurality of directories, and wherein the controller stores the units of data in a same directory as other units of data written to the at least one storage system proximate in time.

47. The content addressable storage system of claim 41, wherein the index is a database table, and wherein the controller stores the one of the units of data in a physical storage location proximate to the physical storage locations of other units of data written to the at least one storage system proximate in time to the one of the units of data.